DOCUMENT RESUME

ED 292 830 TM 011 208

AUTHOR Payne, David A.; Hulme, Gale
TITLE The Development, Pilot Implementation, and Formative Evaluation of a "Grass Roots" Teacher Evaluation System--or--the Search for a Better Lawnmower.
PUB DATE 16 Oct 87
NOTE 20p.; Paper presented at the Annual Meeting of the American Evaluation Association (Boston, MA, October 14-17, 1987).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Counselor Evaluation; Educational Assessment; Elementary Secondary Education; *Formative Evaluation; Media Specialists; Pilot Projects; Principals; *Program Evaluation; Program Implementation; *Teacher Evaluation; Teacher Response
IDENTIFIERS Georgia (Atlanta)

ABSTRACT

The development, pilot implementation, and formative evaluation of a teacher evaluation system for schools in a suburb of Atlanta, Georgia, are described. The system also assessed counselors and media personnel. A committee of 17 teachers and 9 administrative personnel developed the evaluation procedures and instrumentation. A 3-year cycle, which included orientation, assessment, and evaluation phases, was instituted for each teacher. Four schools were involved in the pilot project. Instruments used included an administrator activity log for principals and selected others, a teacher assessment instrument, a teacher survey, and teacher interviews. Results provide insights into changes needed in teacher evaluation procedures and implementation, the impact of teacher evaluation, and teachers' reactions to the evaluation and suggestions for improvement. The evaluation resulted in the following actions by the school district's superintendent: (1) elimination of an unpopular teacher evaluation scale from the design; (2) use of the basic evaluation instrument, which featured eight competencies, as a basis for individualized goal setting; (3) adjustment of the teacher/helper ratio; and (4) continued development of a generic teaching model. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED292830

THE DEVELOPMENT, PILOT IMPLEMENTATION, AND FORMATIVE EVALUATION OF A "GRASS ROOTS" TEACHER EVALUATION SYSTEM - OR - THE SEARCH FOR A BETTER LAWNMOWER

David A. Payne, Director
Educational Research Laboratory
University of Georgia
2085 S. Milledge Avenue
Athens, Georgia 30605
(404) 542-5230

Gale Hulme, Director
Staff Development
Gwinnett Public Schools
Lawrenceville, Georgia 30245

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

David Allen Payne

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 011 208

Paper presented at annual meeting of American Evaluation Association, October 16, 1987 in Boston, MA.

THE DEVELOPMENT, PILOT IMPLEMENTATON, AND FORMATIVE EVALUATION OF A "GRASS ROOTS" TEACHER EVALUATION SYSTEM - OR - THE SEARCH FOR A BETTER LAWNMOWER

The design, development, implementation, and evaluation of any innovation intended for application on a public school system-wide basis is usually fought with frustration, foibles, fizzles, and sometimes fiascoes. This is definitely the case when the innovation is a new teacher evaluation system. Teacher evaluation is a powerful tool that can result in significant improvement in student learning and school climate. If managed poorly, however, it can lead to devisiness, increased anxiety and "evaluation-fear", and possibly the destruction of teacher morale. The evaluation of a new teacher evaluation system, therefore, provides a tremendous opportunity to generate data for formative applications aimed at improvement and the medication of instructional ills..

THE GRASS-ROOTS TEACHER EVALUATION SYSTEM

Authorities have identified several teacher evaluation systems (McGreal, 1983; Darling-Hammond, Wise and Pease, 1983). These range from the highly structured (Medley, Coker, and Soar, 1984) to the artistic and almost mystical (Eisner, 1982). The system described here was developed from a clinical supervision perspective. It emphasized the following activities:

Pre-observation Conference

Observation of Teaching (short and extended)

Feedback and Analysis

Goal Setting

Observation of Teaching

Post-observation Conference and Evaluation

The term "grass-roots" is used here, advisedly, as the design, development, and implementation of the system was a total effort wherein all system educators were represented and/or had direct input. The intent was to develop a system which would meet the following purposes (a) accountability, (b) improvement of instructional effectiveness, (c) encouragement of professional growths, (d) collaboration, (e) planning, and (f) corroboration of employment decisions.

A committee of 17 teachers and 9 administrative personnel developed the evaluation procedures and instrumentation. The total evaluation system included assessments of counselors and media personnel in addition to teachers. Only data on teachers will be presented in this report. The system involved a three-year cycle for each teacher which included orientation, assessment, and evaluation phases. The assessment phase included both long and short term classroom observations. The evaluation phase was only for end-of-cycle teachers.

The pilot implementation also involved (a) workshops with leadership personnel, particularly principals, aimed at enhancing conferencing and observation skills, (b) the refinement of a

generic teaching model based on teacher competencies, (c) publication of a newsletter for teachers (KITE) - Keeping Informed on Teacher Evaluation, and (d) central office meetings with outside consultants to refine the system. Teachers could develop goal plans for the year and present data from a variety of sources to support their performance evaluations. The major theme of the system was, "Improvement through both formal and informal staff development." The system was high-inference, and judgmental as suggested by Popham (1987).

THE SETTING

The pilot project took place in a fast growing southern community (bedroom for Atlanta) where (a) student enrollment was almost 50,000, (b) there were almost 3,000 teachers on staff, and (c) the per pupil expenditure was $2,458 a year. Four schools were involved in the implementation: an elementary (n = 83), middle (n = 60), high (n = 60), and vocational school (n = 11) with a total of 214 teachers.

INSTRUMENTATION

The following are considered to be the psychometric lawnmowers used to trim what had evolved from the grass roots. Administrator Activity Log. Each principal, assistant principal(s), and where relevant leader teachers, were requested to maintain daily logs of their relevant activities and the amount of time spent in each activity. The logs were summarized

weekly over four seven-week blocks. Content analyses of the logs were undertaken and fedback to principals.

Teacher Assessment Instrument. Teachers and principals responded to an eight scale summary instrument in October and again in May. Each scale represented a critical teacher activity. The eight scales were as follows: Knowledge of Subject, Planning, Implementing, Evaluating, Classroom Management, Professional Growth, Professional Responsibilities, and Interpersonal Skills. Judgments were made using four categories: Exceeds Expectations (E), Meets Expectations (M), Needs Improvement (N), and Unsatisfactory (U). Although global judgments were being made each scale had two or more specific indicators to aid the evaluators in synthesizing their judgments (e.g Implements activities in a logical sequence). No performance standards were specified for the evaluation because of the formative nature of this pilot implementation.

Teacher Survey. In as much as pre-project evaluation data might have sensitized the teachers to the innovation, a 30 item retrospective survey form was developed and administered at the end of the school year (Rippey, Seller, and King 1978). The response scale was Better This Year, No Difference, and Better Last Year. Following are two sample items:

The amount of anxiety I feel about being evaluated.

My involvement in the evaluation process.

Teacher Interviews. In an effort to triangulate on teacher perceptiveness of the effectiveness and efficiency of the systems, four teachers were selected at random from each of the pilot schools and interviewed with a semi-structured questionnaire. The content of this questionnaire was derived from the Teacher Survey. Five general questions guided the interviewers (non-pilot teachers) after a session about interview techniques.

RESULTS

Evaluation Question One: What Changes Need to be Made in the Procedures and Implementations?

Initial content analyses of administrator logs yielded four categories: Activity, Reactions, Concerns, and Suggestions. The amount of time associated with each activity was tallied for each team member in each school. It was hoped that these data would reveal how the implementation of the new teacher evaluation system impacted on the activities of, tasks of, and demands made on personnel charged with operationalizing the system. Table 1 contains a summary of the activity data in terms of average number of hours per week for each of the four quarters.

INSERT TABLE 1

ABOUT HERE

The per person averages are based only on the number of individuals actually reporting data for a particular activity. In the interest of brevity only the eight most time consuming activities are reported.

It is interesting to note how the major activity changes from the first period to the last period. At the outset large amounts of time are given over to meetings with central office personnel to work on issues related to implementation of the system and how data collection requirements for the evaluation were to be met. During the second period administrators were involved with making teacher classroom observations for assessment purposes. The last two periods reflect the end product of the process, namely; teacher conferencing for purposes of communicating evaluations. It is also obvious that the aggregate amount of time involved is very large. It, in fact, works out that the three major activities contributing to implementing the evaluation system (Teacher Orientation, Observation, and Teacher Conferences) required an aggregate average of almost 20 hours per week. No meaningful differences were noted between the four levels of schools. The only trend was, as one would expect, that as the number of faculty increase so do time demands. The increase was geometric rather than linear.

Content analyses of the Reactions, Concerns, and Suggestions basically followed the chronology of the implementation.

Evaluation Question Two: What is the Impact of the Evaluation System or Communication Between Teacher and Evaluator?

Percent agreement in the use of the four evaluation categories for the October and May data points is summarized in Table 2. The overall percent agreement for October was 57 and in

---

INSERT TABLE 2

ABOUT HERE

May increased to 65. Although not dramatic the change was in the hypothesized direction. The largest single change for a competency was for Instructional Techniques-Implementing where the input of principal observation data probably had greatest impact.

Analyses of the principal and teacher use of each of the four evaluation categories yielded some interesting results. In the Fall data the contribution to the overall 57% agreement came from 14% of the E category and 43% from the M categories. In the Spring the proportion changed to 25% for E and 40% for M. There

was no contribution from the Needs Improvement and Unsatisfactory classifications.

Not unexpectedly teachers tended to evaluate themselves more favorably than did the principals at both data points. If the four categories are quantified and averaged (E = 4, M = 3, etc.) the following picture of means emerges:

	October	May
Teacher Self-Rating	3.46	3.53
Principal Rating	3.17	3.29

These data suggest an average increase in the evaluations from both groups as well as a decrease in the differences between the group means across time. The convergence is interpreted as enhanced communication between principal and teacher.

Evaluation Question Three: How Do Teachers Evaluate the Evaluation in Process?

Item analyses of the Teacher Survey form led to the elimination of four of the original 30 items. The survey had a Kuder-Richardson internal consistency reliability estimates of .98. The responses (This Year, No Difference, Last Year) were converted to ratings of 3, 2, and 1 and averaged. The mean Teacher Survey score was 62.93 (S = 11.37). This mean expressed as a percent of the maximum possible is 81%. This statistic is

interpreted as supporting this year's evaluation over last year's evaluation procedures.

Responses to individual Teacher Survey items added special insights into teacher opinions. The following three items were highest rated in terms of the "Better This Year" rating.

The Extent of My Input Into the Evaluation Process (64%).

The Extent to Which I Was Able to Share Feelings With My Supervisor About My Job (60%).

The Forms Used to Summarize My Teaching Evaluation (77%).

It is obvious from an examination of the first two items that an important contribution of the new system was to provide the teacher greater active involvement and participation in the overall evaluation process. Teacher "ownership" will obviously enhance the likelihood that the system will be institutionalized. This conclusion is confirmed by qualitative data gathered from interviews. With regard to the evaluation form an apparent conflict exists. Survey data indicate that over-all the teachers liked the form, but interviewer data suggest that the use of the Exceeds Expectations, Meets Expectations, Needs Improvement, and Unsatisfactory evaluative were disliked.

Evaluation Question Four: What Suggestions do Teachers Have for Improving the System?

Five open-ended question probes were used to interview sixteen teachers. They were interviewed by teachers not members

of their faculty. Following is a summary of this free-response data. Although somewhat lengthy, it does capture the flavor of the recorded teacher perceptions.

1. Desecribe The Usefulness Of The Evaluation In Helping You Do A Better Job.

   Almost all teachers were positive. They noted that the evaluation provided explicit objectives, important criteria, and structure for immediate feedback, teacher organization, and more frequent visitation. Great value was seen in providing reinforcement, confirmation, and positive input. It also provided greater self-awareness, and was a great improvement over the old checklist.

2. How Do You Feel About The Ratings E, M. I, U?

   This was the area for greatest concern. Most teachers felt the rating scale was too subjective and had great potential to vary according to each evaluator's interpretation. If used for career advancement, it needs to be clarified. Does the E mean exemplary and thus rare or very effective? The M covers too great a range - from almost excellent to minimal but OK. The improvement process for the I is too inflexible. Some suggested simply S/U with comments, a 1-5 system, or just the dialogue. One

wanted a day to reflect on the rating before signing the evaluation form after conferencing with the principal.

3. To What Extent Did The Evaluation Experience Help You Look At The Total Teaching Process?

Most teachers were positive, noting that the process made them more conscious of their own teaching and provided well-rounded descriptions of the most important areas of teaching. For some the process helped clarify important criteria and tied the whole process of teaching together. Several stressed that it encouraged increased dialogue between faculty and administration and amongst teachers.

Many teachers felt that it didn't substantially change what they did. Weaknesses were noted in that too much time was required of evaluators if they really were to do an effective job. A special education teacher noted that there was a great discrepancy between the teaching model assumed by the instrument and her actual job duties.

4. How Much Confidence Do You Have That Your Supervisor Helped You Improve As A Teacher?

Most were positive, saying that the criticism was helpful because it was constructive and that positively

phrased comments increased their own self-confidence, making them want to continually improve. Comments and dialogue were more helpful than letter ratings. Several said that they had great respect for their evaluator because observations were tailored to the individual; others said increased frequency of visitations added validity to the evaluations.

Several teachers said they had confidence in their principal but that his evaluation was not responsible for their improvement. Concerns were expressed at the secondary level that although they had high ratings their confidence in the evaluation would be strengthened if the department head's input was utilized. They noted that department heads might need training in supervision but that their subject area expertise was very important. A few teachers said that they didn't hear enough of what they were doing well. Several expressed concern that the evaluation process relied heavily on the fairness and competence of the evaluators, and that as the process spread would all be as qualified as this year's group? Several also expressed concern and confusion as to the role of evaluation of both assistant principals and counselor.

Especially concerning the latter, would her role change since she now serves as an administrator?

5. To What Degree Did Being Evaluated Help You Set Goals For Your Teaching?

Positive and negative comments were balanced. On the positive side the process was helpful in giving feedback on whether or not goals were met. Some said it gave structure for their own personal inventory and writing formalized goals kept them on track. Others said the seven areas provided implicit goals.

Many teachers said they didn't set formal goals. Some felt uncomfortable because in their competitive school situation they felt obliged to set goals; that meant they weren't truly optional. A few felt concern that it was unfair that the first time they heard of a weakness was during a formal evaluation. If they had been observed first without judgment they could have set goals to correct weaknesses and that way the negative evaluation wouldn't have gone into their permanent record.

Although these are limited data they do reflect a positive impact of the program, particularly when taken in concert with the quantitative data previously presented. It is obvious that the processes of supervision and evaluation need not be

irreconcilable as suggested by data from McCarty, Kaufman, and Stafford (1986). If the appropriate balance is struck between the gathering of data relevant for decision-making and that for staff improvement a truly valuable evaluation experience can be had by all.

EPILOGUE

So often an external evaluator presents his findings, conclusions, and recommendations to a client and then hears no more about the project. It was gratifying in the present case to find that four significant actions were taken by the superintendent and central office staff as a result of the evaluation. They are as follows:

1. Due to the fact that 50% of the teacher evaluation scale was not being used and that interview data suggested a strong dislike for the scale the rating dimension (E, M. I, and U) was eliminated from the instrument.

2. The basic evaluation instrument with its eight competencies and total of 38 indicators was retained but will be used as a basis for individualized goal setting via a professional development plan.

3. Teacher evaluation is obviously a labor intensive activity (see Evaluation Question One). The data of the present study influenced school leadership personnel to establish a 1:15 supervisor to teacher ratio with the inclusion of peer helpers.

4. Efforts are being increased to refine a generic teaching model tied to the operational objectives-driven curriculum.

A wise evaluator once said, "Reap as you have sown." In the present harvest the reaping was not too grim (and that's no fairy tale), but a more verdent product might have been gathered if better lawnmowers could have been found or created. From the initial seeding came interesting and promising growths, but as the grass grows so do the weeds. It is frequently difficult to separate one from the other. One must be careful not to fertilize incorrectly (or over-fertilize or mis-fertilize) as the seeding may be of discontent rather than enthusiasm. This low-budget evaluation was only partially responsive to Stufflebeams STANDARDS. Lack of time and resources did not allow for the development of maximally responsive instrumentation. For the lack of a good lawnmower, too much grass was lost!

REFERENCES

Darling-Hammond, L., Wise, A. E., and Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. Review of Educational Research, 53(3), 285-328.

Eisner, E. (1982). An artistic approach to supervision. In T. Sergiovanni, Supervision of Teaching: 1982 ASCD Yearbook. Alexandria, VA: Association for Supervision and Curriculum Development.

McCarty, D. J. Kaufman, J. W., and Stafford, J. C. (1986). Supervision and evaluation: Two irreconcilable processes? The Clearing House, 59(April), 351-353.

McGreal, T. L. (1983). Successful teacher evaluation, Alexandria, VA: Association for Supervision and Curriculum Development.

Medley, D. M., Coker, H., and Soar, R. S. (1984). Measurement-based evaluation of teacher performance. NY: Longman.

Popham, J. W. (1987). The shortcomings of champagne teacher evaluations. Journal of Personnel Evaluation in Education, 1(1), 25-28.

Rippey, R., Geller, L. M., and King, D. W. (1978). Retrospective pretesting in the cognitive domain. Evaluation Quarterly, 2(3), 481-491.

TABLE 1

Summary of Result/s of Content Analyses
of Administrator Logs

	Activity	Per Person Average by Quarter (Hours x week)			
		1	2	3	4
1.	Meet With Leadership Team	7	2	3	--
2.	Meet With Central Office Staff	9	7	5	5
3.	Teacher Orientation	5	2	3	--
4.	Observation	6	11	6	7
5.	Teacher Conferences	4	9	9	15
6.	Presentation to Peers	2	2	7	--
7.	Paperwork	4	6	6	8
8.	Individual Work	2	3	2	--

TABLE 2

Percent Agreement Between Principal and Teacher Evaluations for October and May Data Points

Teaching Competency	% Agreement October	May
Knowledge of Subject	67	69
Instructional Techniques-Planning	61	62
Instructional Techniques-Implementing	46	75
Instructional Techniques-Evaluating	60	75
Classroom Management	63	61
Professional Growth	48	66
Professional Responsibilities	57	57
Interpersonal Skills	53	56
TOTAL	57	65